From man to mice - from mice to man

Guest Editorial

From man to mice – from mice to man

We have reached a new landmark in mouse genetics, functional genome annotation and comparative genomics. The sequence and analysis of more than 95% of the mouse genome has been published for the first time in Nature (Vol. 420, 5th December 2002). The biomedical research community has now this superb and freely available tool at their fingertips to tackle the complexities of human disease by using the mouse as a model.

After the publication of the human genome draft sequence a year ago, progress in sequencing has meanwhile covered approximately 98% of the human genome with about 95% in finished form. This resource has now been used for comparisons with the mouse draft sequence. The results provide new important insights into structure and contents of both genomes, with some unexpected surprises. The most important findings are:

- Approximately 99% of all mouse genes have detectable homologues in the human genome.
- The human and mouse gene catalogue contains about 30,000 protein-coding genes, supporting the earlier gene number estimates made from the human draft sequence. But the mouse genome is 14 % smaller than the human genome (2.5 Gb compared with 2.9 Gb).
- About 1,200 new genes have been detected in the human genome by comparisons of mouse and human sequences. Approximately 9,000 genes have been found in the mouse for the first time.
- The analysis of the mouse transcriptome by the FANTOM consortium of the RIKEN Institute has revealed an unexpectedly large number of non-coding RNAs, indicating that non-coding RNA is the major component of the transcriptome.
- Some gene families have experienced a massive expansion in the mouse lineage since the divergence of rodents and humans. Mice have more genes in gene families involved in olfaction, reproduction, immunity and host defence, reflecting different selection forces for these biological functions in the two species.
- By comparing the draft sequence of the inbred strain C57BL/6J with sequences generated from other inbred mouse strains, about 80.000 single nucleotide polymorphisms (SNPs) have been identified. These SNPs are not evenly distributed over the genome. They occur in large blocks, indicating that the founder populations of laboratory mouse strains were extremely limited. The conclusion is that the genetic variation among inbred strains is much smaller than expected.

How will this sequencing endeavour change our daily work? What impact will the mouse genome sequence have on the elucidation of vertebrate gene functions?

Impact on gene targeting experiments

The key to understanding the function of a gene is mutagenesis. By using knockout approaches through homologous recombination in embryonic stem (ES) cells, thousands of medically important loss-of-function mutations in mice have been generated. One of the most time-consuming steps in the knockout business is still the design and cloning of the targeting construct. With the mouse sequence available, tedious restriction site mapping can be avoided. Targeting vectors can be designed from the locus sequence in silico. Thus, the process of generating transgenic mice will be sped-up tremendously. Moreover, sequence annotated C57BL/6J-BAC clones can be directly ordered at genome resource centers and used as DNA cloning resources for targeting constructs. In combination with C57BL/6J-ES cell lines knockout and knockin alleles can be engineered from the beginning in a homogeneous inbred background and will render further extensive backcrosses dispensable.

Impact on large-scale mutagenesis projects

Phenotype-driven large-scale mutagenesis projects have opened a new avenue in the systematic production of new mouse mutants. At least ten big mouse mutagenesis centers have embarked on random chemical mutagenesis by using ethylnitrosourea (ENU) as an effective mutagen. Hundreds of novel mouse mutant lines have been recovered by worldwide mutagenesis screens. All these mutants are now

 0173-0835/02/0601-0067 $17.50+.50/0

waiting for the positional cloning of the underlying molecular gene lesion. With the complete mouse sequence at hand, every gene in a critical mapping interval will be known, allowing a very effective candidate gene approach to find the mutation.

The same is true for several large-scale gene-trap programmes that are underway to add more mutants to the worldwide collection of mutant lines. The availability of the genome sequence will facilitate the determination of the precise integration site of the "gene trap reporter construct". This allows a better selection of targeted ES cells that can be subsequently used to generate interesting mutants.

Revolutionizing the analysis of quantitative traits

One of the exciting brake-throughs can be expected for the genetic characterization of multigenic and quantitative traits in mice. Most of the medically important human diseases like diabetes, hypertension, autoimmunity disorders, susceptibility to cancer and infectious diseases are believed to be determined by complex interactions of multiple genes. Many of these complex traits can also be observed in crosses between inbred strains of mice that serve as animal models for such disorders. The analysis of SNP distributions among inbred strains has now shown that two thirds of the mouse genome have a very low polymorphism rate, due to ancestral haplotypes which have been inherited since the divergence of mouse substrains. This means that positional cloning initiatives of quantitative trait loci (QTLs) should focus on chromosomal subregions of high SNP rates between parental strains. More than 95% of the total genetic variation will be found in these regions. By making use of linkage disequilibrium between a phenotype and particular ancestral haplotypes, it might be possible to identify a critical region where the causative gene should be located. If QTLs have been mapped in several strain combinations, a comparison of haplotype distributions among the strains should allow to narrow the critical region to an even smaller chromosomal interval.

Tasks for the future

To use the full potential of mouse inbred strains in the genetic dissection of important traits, SNP data have to be generated systematically in all commonly used strains. This is a requirement to explore the fine structure of ancestral haplotypes at a resolution high enough for QTL cloning. In addition, the valuable resource of inbred strains needs to be systematically phenotyped for as many physiologically and medically relevant parameters as possible. The recently started Mouse Phenome project at the Jackson Laboratory is an important new tool to make those kinds of data freely available to the scientific community.

Not every human phenotype or disorder can be exactly replicated in mouse models. The elucidation of both genome sequences will further an improved understanding of observed differences between mice and man. But all comparative studies will depend on the to-and-fro between the two genomes, the guiding principle will be maintained - from man to mice and back again from mice to man.

Andreas Lengeling Braunschweig